IMPROVING THE PERFORMANCE OF A MIXED EXCITATION LPC VOCODER IN ACOUSTIC NOISE

Alan V. McCree and Thomas P. Barnwell III

School of Electrical Engineering
Georgia Institute of Technology
Atlanta, GA 30332, U.S.A.

ABSTRACT

This paper presents a number of improvements to our mixed excitation LPC vocoder. First, we have added more sophisticated frequency shaping of the pulse and noise in the mixture. We use a bandpass filter bank to attain a staircase approximation to any desired noise shape. Voicing strength in each frequency band is controlled by periodicity analysis of both the bandpass filtered speech and the bandpass speech envelope. Second, we have improved our pitch detection algorithm by using separate searches on the LPC residual and the input speech signal. Finally, we have added a fixed pulse shaping filter based on a spectrally flattened synthetic glottal pulse. The improved LPC vocoder performs much better in acoustic background noise, and it produces natural sounding speech in both quiet and noisy environments.

1. INTRODUCTION

Traditional pitch-excited LPC vocoders use either a periodic pulse train or white noise as the excitation for an all pole synthesis filter. These vocoders produce intelligible speech at very low bit rates, but they sometimes sound mechanical or buzzy and they are prone to annoying thumps and tonal noises. Since these problems stem from the inability of a simple pulse train to reproduce all kinds of voiced speech, vocoders have been proposed with mixtures of pulse and noise excitation [1, 2]. Mixture excitations are commonly used in formant synthesizers [3, 4] and have also been applied in the context of sinusoidal coding [5]. We have previously developed a mixed excitation LPC vocoder which allows both aperiodic pulses and a pulse/noise mixture [6]. This synthesizer produces more natural sounding speech since it can mimic a richer ensemble of possible speech characteristics.

Our existing mixed excitation vocoder uses a mixture of lowpass filtered pulse train and highpass filtered noise, with the mixture strength controlled by an analysis of the output speech power level. Both the synthesizer model and the control algorithm are intended to model breathy human speech, where the noise has a highpass "whisper" spectrum. As a result, this vocoder performs best for clean input speech. When speech is recorded in an acoustically noisy environment, either voiced speech or background noise could be present at any frequency. This paper describes the modification of our mixed excitation LPC vocoder to allow more complicated spectral shaping of the pulse and noise components in the excitation signal. We also discuss improvements to the pitch detection algorithm and the addition of a pulse shaping filter for better speech quality in acoustic background noise.

2. BANDPASS SYNTHESIZER STRUCTURE

We have developed a synthesizer structure which can generate an excitation signal with different mixtures of pulse and noise in each of a large number (4–10) of frequency bands. There are two equivalent ways to implement this structure, and each viewpoint gives a different perspective. First, this structure can be implemented with a bandpass filter bank. In each band, the pulse and noise are combined based on the voicing strength for that band and bandpass filtered. These filter outputs are then added together to give a fullband excitation signal for LPC synthesis.

Alternatively, this synthesizer can be implemented with only two filters. As shown in Figure 1, the pulse train and noise sequence are each passed through time-varying frequency shaping filters and then added together to give a fullband excitation. For each frame, the pulse filter coefficients are calculated as the sum of each of the bandpass filters weighted by the voicing strength in that band. The noise filter is generated by a similar weighted sum. These filters can give a staircase approximation to any desired frequency shaping.

For ideal bandpass filters, the excitation signal generated by either of these approaches will have a flat power spectrum as long as the sum of the pulse and noise power in each frequency band is kept constant. The important parameters in a practical filter design are the passband and stopband ripple and the amount of pulse distortion. We have chosen to implement our filter bank with FIR filters designed by windowing the ideal bandpass filter impulse responses with a Hamming window. This design technique yields linear phase FIR filters with good frequency response characteristics and the additional benefit of a form of "perfect reconstruction": the sum of all the bandpass filter responses is a digital impulse. Therefore, if all bands are fully voiced, the fullband excitation will be an undistorted pulse. Figure 2 shows the frequency responses of a nonuniform five band design.

3. BANDPASS VOICING ANALYSIS

In order to make full use of this new synthesizer, we need to accurately estimate the degree of voicing in each frequency

0-7803-0532-9/92 $3.00

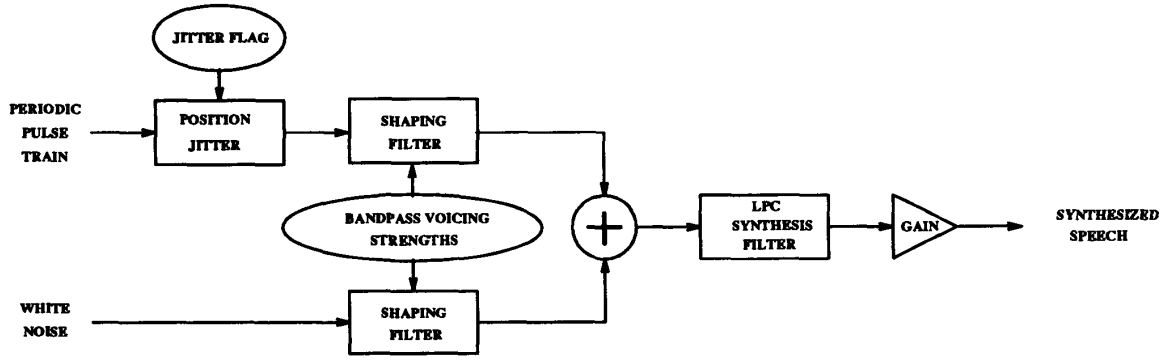

Figure 1: New mixed excitation synthesizer

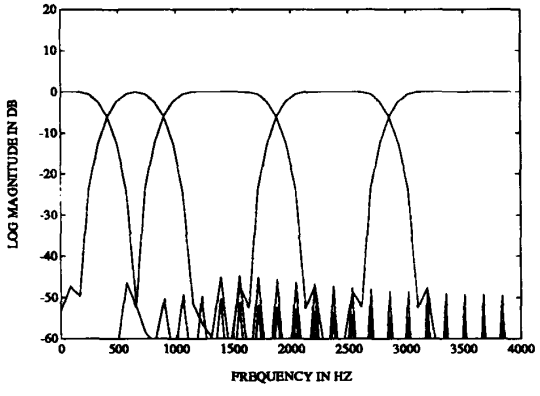

Figure 2: Example bandpass filter responses, 5 band, 48th order.

band. Our old algorithm based on output speech power can generate the frequency dependent voicing decisions needed to control this bandpass synthesizer, but this method assumes a single overall spectral shape of the noise and is not reliable in the presence of acoustic noise.

We have developed an algorithm to estimate the voicing strength in each frequency band by combining two methods of analysis of the bandpass filtered input speech. First, we estimate the periodicity in each band using the strength of normalized autocorrelations around the pitch lag. This technique works well for stationary speech, but the correlation values can be too low in regions of varying pitch. The problem is worst at higher frequencies, and results in a slightly whispered quality to the synthetic speech. Our second algorithm uses a technique similar to time domain analysis of the wideband spectrogram to estimate the voicing strength. The envelopes of the bandpass filtered speech are generated by full wave rectification and lowpass filtering, with a notch filter to remove the DC term from the output. At higher frequencies, these envelopes can be seen to rise and fall with each pitch pulse, just as in the wideband spectrogram. Autocorrelation analysis of these bandpass filter envelopes yields an estimate of the amount of pitch periodicity. Since the peaks in the envelope signal are quite broad, the small pitch fluctuations often encountered in natural speech have little effect on the correlation values. Figure 3 shows examples of these waveforms during an interval of changing pitch. The overall voicing strength in each frequency band is chosen as the largest of the correlation of the bandpass filtered input speech and the correlation of the envelope of the bandpass filtered speech.

4. PITCH DETECTION

Our existing vocoder estimates the pitch period by choosing the largest normalized autocorrelation value of the lowpass filtered LPC residual signal over the range of possible pitch lags. By removing the formant structure, the LPC inverse filter prevents the common problem of finding a higher harmonic of the pitch fundamental. An explicit check for pitch doubling is also included to compensate for the problem of finding slightly stronger periodicity at a multiple of the true pitch period. Finally, gross pitch errors are corrected based on comparison to one past and one future frame.

This algorithm works well for clean input speech, but in a broadband acoustic background noise the residual signal may not contain enough perioidicity. Since the cleanest speech signal is in the formant regions, the LPC inverse filter actually lowers the overall signal to noise ratio by reducing the speech energy while boosting the noise in between the spectral peaks. Therefore, we have added a second pitch search to our algorithm. If the autocorrelation search of the lowpass residual does not give a useful pitch estimate, then a search of the lowpass filtered speech signal is also performed. Since this second search may be influenced by formant structure and find a higher pitch harmonic, the multiple of this estimated pitch period which most closely corresponds to the average value is used. The average value is determined only by recent results from the more reliable residual signal search. This improved algorithm gives better pitch estimates in frames which have poor signal to noise ratio without disturbing the performance for clean speech frames.

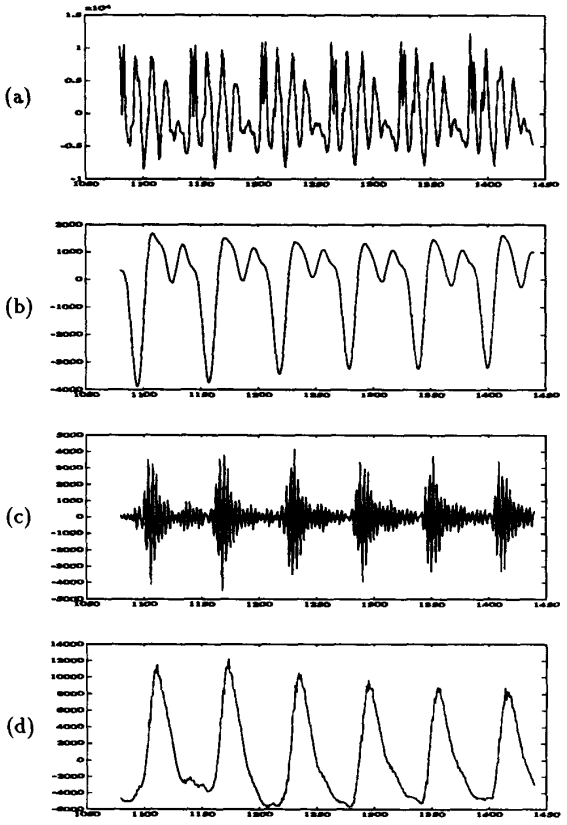

Figure 3: Example speech waveforms for vowel spoken by a male speaker during a region of changing pitch: (a) input speech, (b) bandpass speech, 0–500 Hz, (c) bandpass speech, 2000–3000 Hz, (d) bandpass envelope, 2000–3000 Hz.

5. PULSE SHAPING

This new LPC vocoder largely avoids major artifacts such as buzz, thumps, and tonal noises, but the processed speech still has a slightly unnatural quality. The output sounds somewhat harsh, and acoustic background noise often increases the problem. In comparing bandpass filtered envelopes of input and processed speech, we have noticed some differences in waveforms. Sometimes both waveforms are clearly voiced, but the LPC speech has a more pronounced difference between peak and valley levels. As we mentioned in our previous work [6], the buzz associated with typical LPC vocoders appears to come from peakiness in the higher frequency bands where there should be none. We believe the unnatural quality we discuss here is related to buzz but has two significant differences. First, it stems from a more subtle mismatch in amount of peakiness. Second, it seems to be more significant in the 500–1500 Hz range, rather than at higher frequencies above 2 kHz.

We have seen three different causes for this peakiness mismatch. In some cases, the problem is that there is a higher floor in the bandpass envelope signal of the natural speech due to the background noise level. In other instances, the signal decays too quickly in a frequency band containing a formant because the LPC pole bandwidth is broader than the true formant resonance. Finally, there are times when the natural excitation is not all concentrated at one point in time, perhaps due to a secondary peak from the opening of the glottis or to incomplete glottal closure [7]. These additional excitation points prevent the natural waveform envelope from falling as low as the synthetic signal.

This excessive bandpass waveform peakiness can be removed in a number of ways. The phase coherency between adjacent harmonics can be disturbed with a quadratic phase dispersion [8], by reversing the sign of every fourth harmonic [3], with a fixed randomly generated phase jitter, or by using a high-order allpass filter. The peakiness of the synthetic speech can also be diminished by changing the details of the magnitude spectrum, either by introducing a second pulse in the center of the pitch period or with a fixed randomly generated magnitude jitter. These techniques can reduce the unnatural quality of the synthetic speech, but unfortunately they also introduce distortion. In general too much pulse dispersion gives the speech a rattly or reverberant quality, although the distortion from each approach has a different character.

We have had the most success with a spectrally flattened synthetic glottal pulse, which introduces time-domain spread and changes both magnitude and phase characteristics. A triangle pulse [9, 10] is particularly effective. To remove the low pass frequency response, the pulse is spectrally flattened using a Fourier series expansion. Figure 4 shows some properties of this triangle pulse. The pulse has considerable time-domain spread and detail in its magnitude spectrum. Slight variations in pulse shape can significantly affect both these waveform characteristics and the processed speech quality. Using this FIR pulse shaping postfilter decreases the synthetic bandpass waveform peakiness and results in more natural sounding LPC speech output for both clean and noisy speech. Of course, pulse shaping only gives a minor decrease in peakiness, so the mixture excitation is still needed to remove the buzz.

6. IMPLEMENTATION AND EVALUATION

We have added these features to our existing implementation of a mixed excitation LPC vocoder, which runs in real-time on a personal computer with plug-in boards based on the TMS320C30 DSP chip. The new vocoder runs at 3000 bps with the bit allocation shown in Table 1. Since the primary purpose of this work is to demonstrate the possiblities of this new vocoder model, the number of bits used for the conventional LPC parameters is generous. Comparable speech quality may well be attainable at bit rates below 2000 bps. In informal comparisons with our previous mixed excitation vocoder, the new coder produces significantly higher quality speech in the presence of acoustic noise, and also gives some improvement for clean input signals.

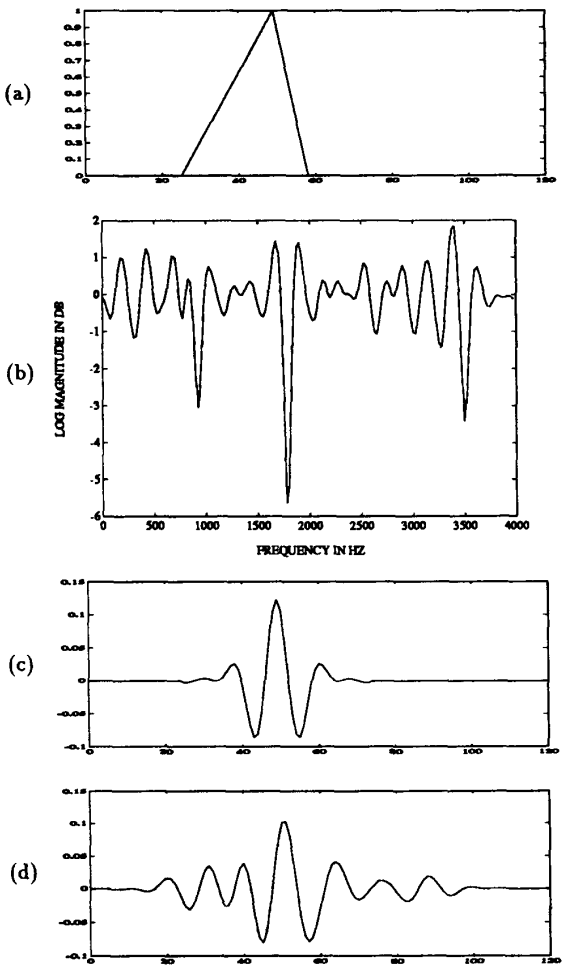

Figure 4: Synthetic triangle pulse: (a) triangle waveform, (b) Fourier transform after spectral flattening, (c) bandpass impulse, 500–1000 Hz, (d) bandpass synthetic pulse, 500–1000 Hz.

7. CONCLUSION

This paper has presented improvements to our mixed excitation LPC vocoder. The addition of a more sophisticated bandpass filter synthesizer model and bandpass voicing analysis combine to greatly reduce the buzzy quality typically associated with LPC vocoders in broadband acoustic background noise. Also, a pitch detection algorithm using both the LPC residual signal and the input speech can produce better pitch estimates in a noisy environment. Finally, a pulse shaping postfilter can significantly improve the naturalness of the LPC speech output for both quiet and noisy inputs.

LPC coefficients (12 LSP's)	43
gain	5
pitch and overall voicing	7
bandpass voicing	5-1
jitter flag (aperiodic pulses)	1
TOTAL: 60 bits / 20 msec = 3000 bps	

Table 1: 3000 bps LPC vocoder bit allocation

REFERENCES

[1] J. Makhoul, R. Viswanathan, R. Schwartz, and A. W. F. Huggins, "A Mixed-Source Model for Speech Compression and Synthesis," *J. Acoust. Soc. Amer.*, vol. 64, pp. 1577–1581, Dec 1978.

[2] S. Y. Kwon and A. J. Goldberg, "An Enhanced LPC Vocoder with no Voiced/Unvoiced Switch," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 32, pp. 851–858, Aug 1984.

[3] J. N. Holmes, "The Influence of Glottal Waveform on the Naturalness of Speech from a Parallel Formant Synthesizer," *IEEE Trans. Audio and Electroacoustics*, vol. 21, pp. 298–305, June 1973.

[4] D. H. Klatt, "Review of Text-to-speech Conversion for English," *J. Acoust. Soc. Amer.*, vol. 82, pp. 737–793, Sep 1987.

[5] D. W. Griffin and J. S. Lim, "Multiband Excitation Vocoder," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1223–1235, Aug 1988.

[6] A. V. McCree and T. P. Barnwell III, "A New Mixed Excitation LPC Vocoder," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 593–596, 1991.

[7] J. N. Holmes, "Formant Excitation Before and After Glottal Closure," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 39–42, 1976.

[8] R. J. McAulay and T. F. Quatieri, "Multirate Sinusoidal Transform Coding at Rates from 2.4 kbps to 4.8 kbps," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 1645–1648, 1987.

[9] M. R. Sambur, A. E. Rosenberg, L. R. Rabiner, and C. A. McGonegal, "On Reducing the Buzz in LPC Synthesis," *J. Acoust. Soc. Amer.*, vol. 63, pp. 918–925, Mar 1978.

[10] A. E. Rosenberg, "Effect of Glottal Pulse Shape on the Quality of Natural Vowels," *J. Acoust. Soc. Amer.*, vol. 49, pp. 583–590, 1971.